THE PHILIPPINE STATISTICIAN
Volume XVII, Nos. 1/2, 1968

FORECASTING CYCLICAL CROP PRODUCTION

by JOSE S. GUTIERREZ[1]

1. **Introduction.** Crop forecasting is difficult because of the large numbers and complex relationship of factors that influence the agricultural systems. Many of these relations are not adaptable to precise and easy forecasting. Many are undergoing continuous changes from preplanting time to post harvest period. Since in appraising current prospects crops reporters take into consideration seasonal progress, pests, diseases, amount of fertilizer used and cultural practices, crop condition reports therefore reflect the complex effect of these factors.

2. **Crop Production Patterns.** A question often asked is "Are crop yield variations random?" or the same as asking "is there really a yield—weather pattern?" Fluctuations in both weather and crop yields, whether short or long range are almost universally looked upon as a matter of chance. Statistical studies have shown that fluctuations of crops and weather are essentially similar to what might be expected in a series of random numbers (Foote and Bean [3]). Bean [2], however, restated his views on the evidence of trends and patterns and suggested this as the first step in determining the reality of such trends and patterns. Foote's test showed that standard statistical tests do not appear to be sensitive in distinguishing between a random series and those made up of repeating patterns.

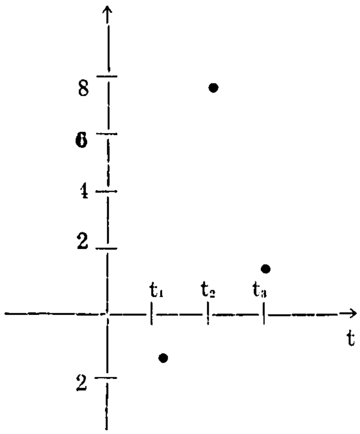
Meteorologists, on the other hand, believe that progress could be made and that the day when season-to-season long range

[1] Assistant Professor, U.P. Statistical Training Center.

weather and crop yield will be possible. If statisticians can show the existence of weather-yield patterns and can determine what they are, this in itself will be a long step toward the solution of crop-weather forecasting (U.S.D.A. 1941 Yearbook of Agriculture). Even if the chance of success is relatively small, all possibilities should be examined inasmuch as even only slight improvements in forecasting could entail great economic benefits. Imperfect crop forecasts are not necessarily valueless. To many people they are useful and considered unbiased. Market manipulations and misinformation are prevented by its use. And it can and should be improved (Baker [1]).

Yield, which is the principal yearly determinant of production, is primarily influenced by weather factors. Morgan (1961) stated that current prospects which reflect the impact of weather and other factors to date as well as thereafter, is reflected in the dependent variable, final yield per acre. This implies that weather and yield forecasting are inseparably linked. Bean (1942) pointed out that there persists between one decade and another, between alternative decades or even in longer intervals the existence of both weather and crop patterns.

The writer [5] noted that rice production in this country seems to follow a 3-year crop production pattern as shown in the following graph:



Over a long period the pattern will look like this:

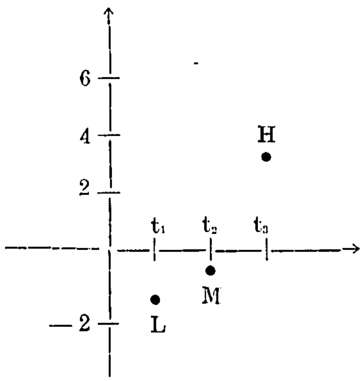
L H M   L H M . . .

The sequence of L, H and M will of course change depending upon the starting time of reference, for example,

H M L   H M L . . . or

M L H   M L H . . .

Another crop which was observed to exhibit a production pattern is coffee. The pattern is as follows:



What is interesting to note in the production patterns of rice and coffee is that the low years coincide. However, it takes coffee a longer time to recover from a bad production year in comparison to that of rice (see the above two graphs). This is plausible since a poor production year is usually the result of drought or inclement weather, while the strain on the root system of a perennial crop requires some time before it can be overcome.

3. **Regression Methods of Crop Forecasting.** The usefulness of a crop forecast depends upon the timelessness of its release. Simplicity of form and the ease of computations are highly priced premiums in the consideration of the equation or system of equations to be adopted. It is for these reasons, perhaps, that regression regression-type estimators are becoming

tools of major importance in the statistical forecasting workshop.

Classical Regression Theory. In this situation we have a set of n values of y, say, y_j ($j = 1, 2, \ldots, n$) corresponding to values of x_i, say, x_{ij} ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$) and we postulate

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \ldots & x_{1n} \\ \ldots & \ldots & \ldots \\ x_{n1} & \ldots & x_{nn} \end{bmatrix} \begin{bmatrix} \beta_n \\ \vdots \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or in matrix form $y = X\beta + \epsilon.$

The assumptions under the classical regression theory are:

1. $E(\epsilon) = O$,

2. $V(\epsilon) = E(\epsilon\epsilon') = \sigma^2 I_n$ (the homoscedastic assumption),

3. The x_{ij}'s are given, i. e. they are fixed in repeated samples. (As an alternative to this assumption, it is sometimes assumed that the x_{ij}'s are random variables with joint distribution independent of the ϵ's).

4. Either the rank of X is $k \leqq n$ or $X'X$ is non-singular and $\beta = \mathfrak{b}$ is a unique solution of the set of normal equations $X'X\beta = X'y$. The variance of β in the second case is $\sigma^2 (X'X)^{-1}$ which is an unbiased estimator of β.

Weighted Regression Theory. Suppose that the second assumption in the classical regression theory is not met and there is a non-singular matrix T such that $TVT' = I$. The normal equations under this situation will be

$$X'V^{-1}X\beta = X'V^{-1}y.$$

The variance of β is $\sigma^2 (X'V^{-1}X)^{-1}$ in this case. This is the so-called classical weighted regression theory. In actual forecasting, the matrix T is not used, but an arbitrary multiplier to give different weights to the x and y variables is often adopted. The weights in this connection depend on the cycle or pattern of crop production; for example, for a three-year cycle the weights can simply be 1, 2, and 3, i.e.

Year	Weight
1	1
2	2
3	3
4	1
5	2
6	3
7	1
8	2
9	3

If the years are the x's, then the above scheme will yield the following values for x:

$$1, 4, 9; 4, 10, 18; 7, 16, 9$$

The y's for this procedure may or may not be weighted.

The model in case the x's are weighted will be given by

$$y = {}_{\omega}Z\beta + \epsilon,$$

where ${}_{\omega}Z = X$ in the classical model.

The normal equations will be $({}_{\omega}Z)'({}_{\omega}Z)\beta = ({}_{\omega}Z')y$ and the variance of β is $\sigma^2 [({}_{\omega}Z)'({}_{\omega}Z)]^{-1}$. This approach is basically a classical regression technique except for the introduced multiplier, resulting to non-equidistant x's. However, if the y's have been weighted, i.e.

$${}_{\omega}y = {}_{\omega}Z\beta + {}_{\omega}\epsilon,$$

then the classical weighted regression theory applies.

Grouped Regression Theory. In view of the apparent production pattern of crops in this country, the use of classical regression methods of estimation will yield an over-estimated forecast during poor crop seasons and under-estimated forecasts during good crop seasons. A better approach to crop forecasting is to take the production pattern into consideration, perhaps, more in detail as that in weighted regression.

Houthakker regarded grouping of observations as some kind of a matrix transformation. Suppose we wish to replace y_1, y_2, y_3 by their mean value $\bar{y}_1$, y_4, y_5 by their mean value $\bar{y}_2$. This can be done by the matrix transformation

$$\begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}.$$

More generally, the m groups $[y_1, \ldots, y_{r1}]$, $[y_{y1+1}, \ldots, y_{r1+r2}]$, $\ldots$, $(y_{n-r_m+1}, \ldots, y_n]$ is represented as follows:

$$\begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} = \begin{bmatrix} G_{r1} & \cdots\cdots & 0 \\ 0 & & G_{rm} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

or

$$\bar{y}_G = Gy,$$

where G_{ri} is a row submatrix $(1/r_i \ 1/r_i \ \ldots \ 1/r_i)$ with r_i columns.

It should be noted that grouping of $y_1, \ldots, y_n$ into m groups containing $r_1, \ldots, r_m$ observations and replacing each by their means is effected by the transformation GP where P is a permutation matrix of order n. For example, suppose that instead of grouping $y_1, \ldots, y_5$ into y_1, y_2, y_3 and y_4, y_5, we group them into y_1, y_4, y_5 and y_2, y_3. Then this is effected by

$$\bar{y} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \end{bmatrix} y =$$

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} . y$$

Consider again the classical regression model

$$y = X\beta + \epsilon.$$

Then, by grouping we obtain the following

$$Gy = GX\beta + G\epsilon \text{ or}$$

$$\bar{y} = \bar{x}\beta + \eta,$$

where $\bar{y} = Gy$, $\bar{x} = Gx$, and $\eta = G\epsilon$. The least squares estimate of β is

$$\bar{b} = [\bar{x}'(GG')^{-1}\bar{x}]^{-1}x'(GG')^{-1}\bar{y} = [X'HX]^{-1}X'Hy,$$

where $H = G'(GG')^{-1}G$. The above estimator can readily be shown to be unbiased and with a variance

$$V(\bar{b}) = \sigma^2[\bar{x}'(GG')^{-1}\bar{x}]^{-1} = \sigma^2[X'HX]^{-1}.$$

The efficiency of the method of grouping is to be measured by the variances of b as compared with those of $\bar{b}$, that is, by the diagonal elements of $\sigma^2(X'HX)^{-1}$ as compared with those of $\sigma^2(X'X)^{-1}$. The diagonal elements of $\sigma^2(X'HX)^{-1} \geqq$ the diagonal elements of $\sigma^2(X'X)^{-1}$ are both BLUE (i.e. best linear unbiased estimators) a measure of the efficiency of the method of grouping could be taken to be

$$\frac{\text{tr } (X'X)^{-1}}{\text{tr } (X'HX)^{-1}} \leqq 1.$$

4. **Applications.** The various procedures of obtaining regression equations discussed above were applied to the annual

increase or decrease in Philippine rice production from 1955 to 1965. The data are as follows:

Year	Percent Increment
1955	0.64
1956	2.20
1957	2.22
1958	-4.26
1959	15.01
1960	1.49
1961	-0.93
1962	5.54
1963	1.45
1964	-3.13
1965	3.90

(Source of data: Bureau of Agricultural Economics, DANR)

Classical Regression Model: The estimating equation for this case applied to the above data is as follows:

$$\hat{y}_c = 2.40 - \underset{(0.68)}{0.0334}\, x$$

where (0.16) is the standard error of the regression coefficients for this model.

Weighted Regression Model: The first model obtained using a weighted x is as follows

$$\hat{y}_Q = 3.93 + \underset{(0.68)}{0.1488}\, x$$

and for the classical weighted regression model the estimating equation is given by the following:

$$\hat{y}_c = 3.10 + \underset{(0.4071)}{0.3592}\, x$$

Grouped Regression Model. A simple way to group observations is to first consider the y-values of interest and assign consecutive numbers on the corresponding x's. To illustrate this method consider the following values of x and y:

Year	x	y
1950	1	y_1
1951	2	y_2
1952	3	y_3
1953	4	y_4
1954	5	y_5
1955	6	y_6
1956	7	y_7
1957	8	y_8
1958	9	y_9

Regrouping the y observations and assigning consecutive values to the corresponding x's will yield the following groupings:

Group	Year	x	y
	1950	1	y_1
I	1953	2	y_4
	1956	3	y_7
	1951	1	y_2
II	1954	2	y_5
	1957	3	y_8
	1952	1	y_3
III	1955	2	y_6
	1958	3	y_9

Thus the problem of forecasting becomes a computation of three separated simple linear regression equations, which are respectively:

Group I: $\hat{y}_1 = a_1 + b_1 x,$

Group II: $\hat{y}_2 = a_2 + b_2 x,$

Group III: $\hat{y}_3 = a_3 + b_3 x.$

The other method requires the use of grouping or transformation matrices. For the above data these matrices are:

Group I: $$G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Group II: $$G_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 4 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Group III: $$G_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

When these matrices are applied to the classical regression model, the resulting grouping becomes as follows:

Group	Year	x	y
I	1950	1	y_1
	1953	4	y_4
	1956	7	y_7
II	1951	2	y_2
	1954	5	y_5
	1957	8	y_8
III	1952	3	y_3
	1955	6	y_6
	1958	9	y_9

Applications of the above methods to the annual increment yielded the following models:

Grouping Plan 1:

$$\hat{y}_{11} = -3.92 + \underset{(1.06)}{0.7980}\ x$$

$$\hat{y}_{12} = 7.74 - \underset{(3.14)}{0.4380}\ x$$

$$\hat{y}_{13} = 2.90 - \underset{(0.14)}{0.5923}\ x$$

Grouping Plan 2:

$$\hat{y}_{21} = -3.38 + \underset{(0.33)}{2.660}\ x$$

$$\hat{y}_{22} = 7.37 - \underset{(0.91)}{0.1134}\ x$$

$$\hat{y}_{23} = 2.49 - \underset{(0.20)}{0.1283}\ x$$

For forecasting purposes the equations obtained for Group I will be used for the years 1955 + 3i; Group II, 1956 + 3j; Group III, 1957 + 3k, where i, j = 4, 5, 6, ... and k = 3, 4, 5, 6, ...

Common regression coefficients were computed for Grouping Plan 1 and Grouping Plan 2. The common regression coefficients for both plans are the simple arithmetic mean of the separate regression coefficients for each group for each plan. However, a common regression coefficient for the second plan was computed as a weighted mean of the regression coefficients by group. The weights used were proportional to the sum of the values of the x's used.

The models employed using common regression coefficients are as follows:

Plan 1: $\hat{y}_{1c} = 3.38 - 0.2323\ x.$

Plan 2: $\hat{y}_{21c} = 5.75 - 0.5942\ x.$

$\hat{y}_{22c} = 2.34 + 0.0243\ x.$

Comparison of the Predicted Rate of Increase of Palay Production for 1966. The rate of increase of palay production for 1966 was predicted using the different models obtained above. The results of the aforementioned prediction are as follows:

METHOD	MODEL OR PREDICTING EQUATION	Predicted Rate of Increase of Production for 1966 (standard error)
Classical	$\hat{y}_c = 2.40 - 0.0334\ x$	+ 2.00 (6.42)
Grouping Plan 1		
Common Regression	$\hat{y}_{1c} = 3.38 - 0.2323x$	+ 0.60 (6.41)
Group III	$\hat{y}_{13} = 2.96 - 0.5923x$	+ 0.53 (0.55)
Grouping Plan 2		
Common Regression	$\hat{y}_{22c} = 5.75 - 0.5942x$	−1.38 (6.40)
	$\hat{y}_{22c} = 2.34 + 0.0243x$	+2.63 (6.41)
Group III	$\hat{y}_{23} = 2.49 - 0.1283x$	+0.95 (2.11)
Weighted Regression		
Only x's weighted	$\hat{y}_Q = 3.93 + 0.1488x$	9.28 (5.32)
Both x's and y's weighted	$\hat{y}_w = 3.10 + 0.3592x$	7.28 (4.68)

Easily noticeable from the above table is the relatively larger standard error for the predicted rate of increase of palay production for 1966 using predicting models with common regression coefficients. The predicting equation for the common regression of grouping plan 2 yielded a negative rate of increase of palay production for 1966, while all others yielded positive rates.

The results of these preliminary studies seem to indicate the desirability and feasibility of incorporating in the prediction model the crop production pattern. This crop production pattern can be considered in the formulation of the predicting equation by the use of an appropriate grouping matrix.

5. **Summary and Conclusion**. The desirability and feasibility of using a new approach to crop production forecasts in this country have been demonstrated in this study. The consideration of the production pattern in the formulation of the predicting equation may eliminate overestimated forecasts in years of lean harvests and underestimated forecasts in years of good production. For lack, however, of data for other crops the approach described in this study was applied only to rice. Moreover, this study has been confined to only one form of grouping matrices. A more comprehensive treatment may be necessary.

REFERENCES

[1] BAKER, D. S. "Evaluation of wheat crop production estimates", *Agricultural Economics Research*, Volume 2 (1953) pp. 27-28.

[2] BEAN, L. H. "Crop yields and weather", *U. S. D. A. Agr. Misc. Pub. No. 471* (1955) pp. 1—3.

[3] FOOTE, R. L. and L. H. BEAN. "Are yearly variations in crop yields really random", *Agricultural Economics Research*, Volume 3 (1951) 23—30.

[4] GUTIERREZ, Jose R. "Forecast of rice and corn production in the Philippines", *The Statistical Reporter,* Volume 3 (1960) pp. 14—22.

[5] GUTIERREZ, Jose R. A review: "The System of Evaluating the Rice Situation and Formulating Government Policy in the Philippines", *Papers and Reviews: Rice and Related Statistics—The Present Situation and Outlook.* Manila: Statistical Training Center, 1966.

[6] PRAIS, S. J. and AITCHISON. "'The grouping of observations in regression analysis", *Bulletin de l'Institut International de Statistique* 1958.